HW4 solutions:

Section 30.3.

(3) @. (131) 133 135 138 137 137 139 (140) 141 141 141
143 143 146 (146) 147 147 148 149 149 149 (149)
151 152 153 154 154 155 (164)

Five number summary: 131, 140, 146, 149, 164.

155 would need to substitute 164 since 164 is an outlier.

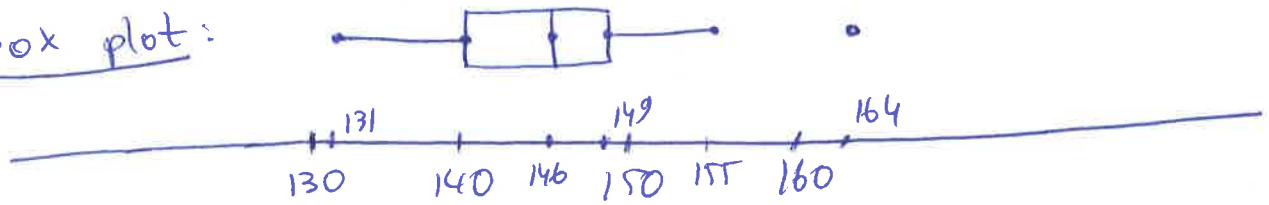
IQR: 149 − 140 = 9.

$140 - 1\frac{1}{2} \times 9 = 140 - 13.5 = 126.5$

$149 + 1\frac{1}{2} \times 9 = 149 + 13.5 = 162.5$.

Since 164 does not lie between 126.5 and 162.5, it is an outlier.

Box plot:



(b) If we use the first and the third quartile scores, we will be missing heights of 50% of the girls.

---

(4) @ 1 1 1 1 1 1 2 2 2 2 2 2

(b) Not possible

(c) 1 1 1 1 1 1 2 2 2 2 2

---

(5) @ 5 5 5 | 5 6 6 7 7 | 9 9 9. 5-number summary: 5, 5, 6, 8, 9.

(b) 5 5 5 | 5 6 6 7 9 | 9 9 9. 5-number summary: 5, 5, 6, 9, 9.

(c) 1 10 10 10 10 10 10 10 10 10 10

(d) 7 7 7 | 7 7 7 7 7 | 8 8 5-number summary: 7, 7, 7, 8.5, 8.

(e) 8 8 8 8 8 8 8 8 8 8 8 8 5-number-summary: 8,8,8,8

(f) For a, b, d, e the underlined numbers in the 5-number-summary are equal and that gives desired condition.

For c, IQR = 0, and 1^{st} quart. $- 1\frac{1}{2} \times IQR = 10$

3^{rd} quart. $+ 1\frac{1}{2} \times IQR = 10$

since 1 is not equal to 10, it is an outlier.

---

(6) (a) It is difficult to tell without knowing exact scores but we can tell that about 50% of students scored above 82%, which is pretty nice.

(b) About 25% of 32 students, or 8.

(c) About 5.

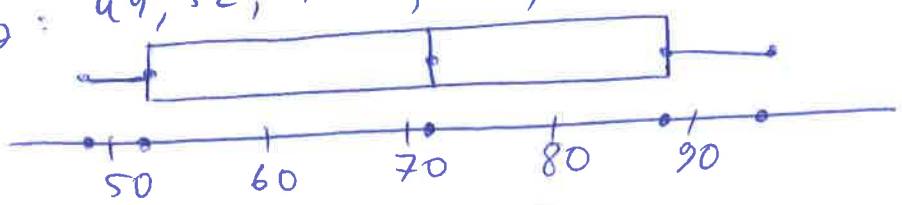
(d) Perhaps score was recorded incorrectly.

SLES

(49) 50 51 51 (52) 56 58 58 62 78 81 82 82 (88) 88 89 92, (93).

5-number-summary: 49, 52, 70.5, 88, 93

a. Box plot:



b. Individual data values.

Section 30.4:

(1) (a) Yes (8, 9, 10, 11, 12)

(b) Yes (8, 8, 9, 11, 14)

(c) Yes (1, 2, 3, 6, 38)

(d) Yes (10, 10, 10, 10, 10)

(e) Yes (1, 2, 3, 6, 38).

(f) No. If one thinks of the mean as balancing point of the data set, then the mean would have to fall between the largest and the smallest values of the data set.

(3) 0, 10, 10, 10, 10.

1st quartile = 10, mean = $\frac{0+10+10+10+10}{5} = 8$. $8 < 10$. ✓

---

(11) $\frac{21 \times 3 + 6 \times 8}{3+8} = \frac{111}{11} = 10.091$.

---

(13) $\frac{64 \times 3 + 86 \times 2 + 110 \times 2}{7} = 83.43$.

SLE2 (a) Mean = 11, median = 7.5

(b) Mean = 8.45, median = 7.

The mean is affected more than the median by the relatively large 39. When the 39 is removed, the median is moved by at most one case.

Section 30.5:

(1) The average deviations for Gonzales = 2 minutes
The average deviations for Childress = 4.4 minutes.

Average deviations provide [redacted] data that cannot be obtained by just looking at the mean — namely, Prof. Gonzales is more uniform in spending time with students, his data is less spread out.

---

(2) (a) B, because of the one high value.

(b) D, because of the very low value at the end.

(c) B or D, but it is hard to tell which, both have high variation.

(d) C, because there is very little variation among the numbers.

SLE1 (a) 2, 5, 8, 9.9, 12, 13, 16, 19, 19, 21, 24.

Range = 24 − 2 = 22.

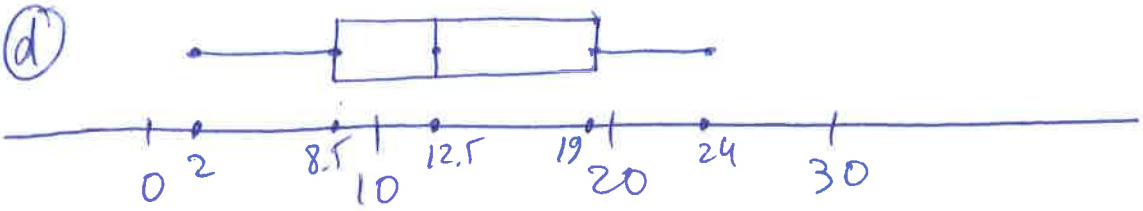
(b) 2, 8.5, 12.5, 19, 24.

(c) $IQR = 19 - 8.5 = 10.5$.

$8.5 - 1\frac{1}{5} \times IQR = -7.25$ $\quad 19 + 1\frac{1}{2} \times IQR = 34.75$.

All data values fall between -7.25 and 34.75, so there are no outliers.

(d)



(e) Mean ≈ 13 (f) Two modes, 9 & 19.

(g) The mean (13) & the median (12.5) are close to each other. Because there is no single mode, we cannot compare mode with the median and the mean.

(h) Average deviations = 5.6.

(i) standard deviation = 6.5.

(j) 5.6 and 6.5 are fairly close. Both numbers provide information about the spread of the data. The five-number summary is also an indication of spread, but of a different type.

Section 31.1:

(3) (a)

	Chen	Lopez	Wilson	Totals
K	7	4	8	19
L	2	7	5	14
R	3	0	1	4
S	3	2	2	7
T	5	10	6	21

You may also consider bar graphs or pie charts for individual classes.

Looks like K is the favorite in two classes and T is the favorite in one class but also pretty popular in the other two classes.

(b) T is most liked among all children in three classes. But K is also popular.

(c) A teacher's enthusiasm might influence children's responses.

④ @ Class A appears to have more high-scoring students than does Class B, although it also has more low-scoring ones as well. Usually in competitions only the highest scores from a team are counted, so Class A would be a better choice.

ⓑ "That would be one way to do it, although Class A has more high scores. Let's find out whether everyone's score counts for the team, or just the highest ones."

---

⑤ @ Rather than quartiles at the ends of the boxes, the ends are at the 20th and 80th percentiles.

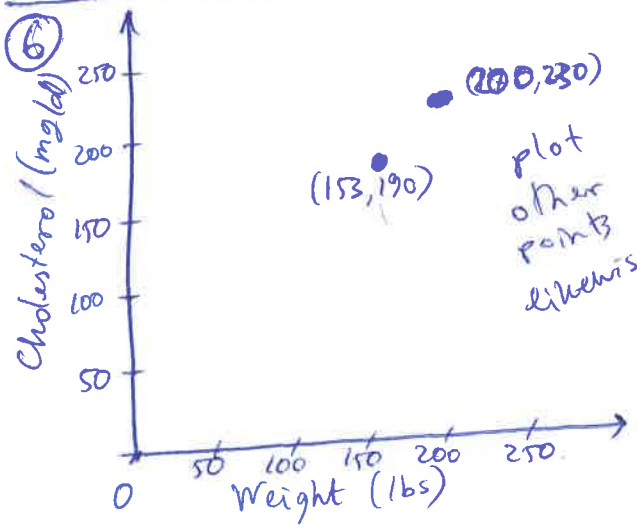
ⓑ Number sense: Males are slightly better, and are not so spread out as females.

Measurement: Females and males perform alike, although the males' lowest 20% are more spread out than those of females.

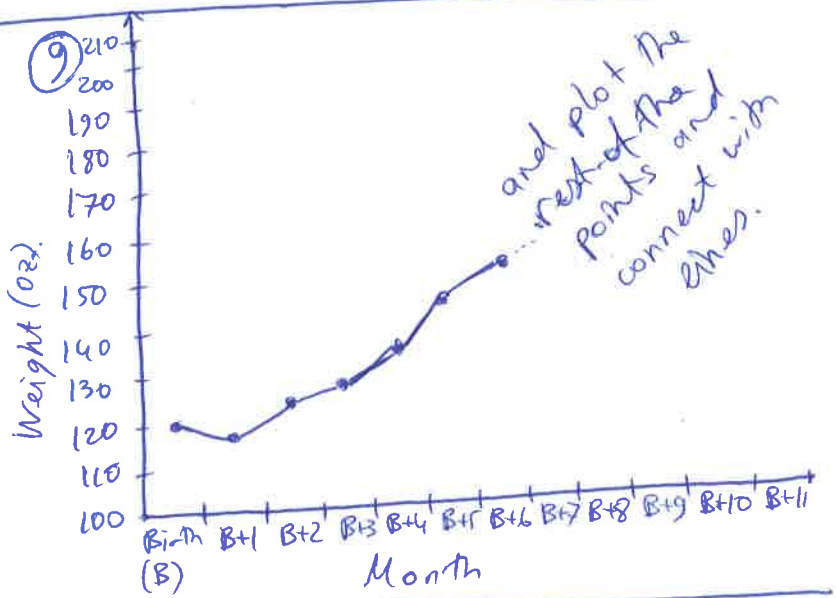
Spatial sense: There are no striking differences, although the middle 50% of the females vary less than the middle 50% of the males.

---

⑥



⑨



Section 31.2:

① @ 1. Older trees have thicker trunks.

ⓑ 0. The temperature inside stays about the same throughout the year

ⓒ -1. Price goes up, the demand goes down.

ⓓ 0. A car can cost more because of high mileage. but many expensive cars also have low mileage.

(e). 0. Having more time on the test will not allow a student to do solve a problem he/she does not know how to solve.

(f). 1. Hopefully spending more time means learning more.

---

(2) (a) Humans, Asian elephants, African elephants.

(b) Such a high positive correlation means that as body weight increases, so will the brain weight.

(c) Humans are at (62, 1320) which is farther away from the line than most other points. So we could say Humans are outliers.

(d) $0.94 \cdot (2000) + 191.22 = 2071.22$ grams.

(e) There is a strong positive correlation between the two weights. This is not surprising since one would expect larger animals to have larger brains.

---

(5) (a) $\sqrt{0.81} = .9$

(b) About 15.9; about 185.

(c) 200

(d) Not necessarily. Correlations do not imply causation.